DOCUMENT RESUME

ED 266 691 FL 015 547

AUTHOR Voss, Bernd
TITLE Language Testing in German Secondary Schools--Theory and Practice.
PUB DATE 85
NOTE 20p.; In: Language Testing in School. AFinLA Yearbook 1985. No. 41 (see FL 015 537). Document is in small print.
PUB TYPE Reports - Descriptive (141)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Administrative Organization; Criterion Referenced Tests; English (Second Language); Foreign Countries; Language Proficiency; *Language Tests; Norm Referenced Tests; Scoring; Secondary Education; Second Language Instruction; Student Evaluation; *Test Construction; *Testing; *Test Use
IDENTIFIERS *West Germany

ABSTRACT

The ways in which achievement in a foreign language is assessed in German secondary schools are outlined and illustrated with examples from the testing of English. The overall organization of assessment procedures is described, the situations for forms 5-10 and 11-13 are differentiated, and the regulations governing conditions for the "Abitur" (pre-university, school-leaving) examination are discussed. Problems in the testing system are examined in some detail; it is noted that the system has the advantage of flexibility and freedom, which allow teachers to tailor their assessment procedures to their teaching methods. It is suggested that the system contains an untapped potential for criterion-referenced testing, and questions are raised about the degree of norm- referencing found in the system. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J Tomala

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED266691

LANGUAGE TESTING IN GERMAN SECONDARY SCHOOLS - THEORY AND PRACTICE

Bernd Voss (University of Bielefeld)

1 Introduction

The purpose of the present paper is to give an outline of the ways in which achievement in foreign languages is assessed within the context of the German school system.

It should be noted that in the above introductory sentence the term 'test' has been rather carefully avoided because - as will become fully apparent later - the assessment procedures used in the Federal Republic of Germany tend to be both less rigorous and more global than 'tests' in the usual understanding of the word (1). If we maintain that tests be objective, reliable, and valid, or that at least serious attempts be made to make them such, then with very few exceptions the procedures by which achievement in foreign languages is assessed within the German school system are not tests. In other words, there is no tradition of testing in the 'hard' sense of the word, in German schools.

This is partly encouraged by the fact that in Germany, as a Federal Republic, education is run not by a single national Department of Education, with resulting tendencies towards uniformity throughout the country, but - since education is constitutionally a matter of the individual states - by 11 independent Education Departments, each organising matters in somewhat different ways, in each of the parts of the country. In what follows we shall take the situation in the most populous state of North-Rhine Westphalia (NRW) as an example in the hope that this will at the same time mirror enough of the common ground that does exist between the approaches of the individual states, without, however, being able to claim correctness of detail for all the other states as well.

After a brief exposition of the organisatory framework of assessment in general, we shall follow the structure of the German secondary school system and deal with forms 5 - 10 and 11 - 13 as two different groups, before finally turning to the procedures used in the final (pre-university level) school-leaving examination, at the highest level of a student's school career. In each case, we shall present the official rules and regulations as issued by the Department of Education, and then try to explain in which way these tend to be implemented by teachers in the classroom, thus aiming at a survey of both the theory and current practice.

2 General Organisatory Framework of Assessment within the School System

Modern languages, as school subjects, operate under conditions

FL015542

2 BEST COPY AVAILABLE

which are in no way different from those of other subjects. The appropriate assessment procedures, accordingly, need to conform to the general regulations that govern the ways in which the results of teaching and learning efforts are captured and measured within the system in general.

For North-Rhine Westphalia, this general framework is laid down in the Allgemeine Schulordnung (General School Regulations) of 1978, which stipulate, in § 21, that all assessment must refer to both knowledge and skills and needs to be based on all aspects of a pupil's work, in particular (i) on written pieces of work as elicited by a prescribed number of Klassenarbeiten (written classroom tests), which carry half the weight of the total assessment, and (ii) on what is called Sonstige Mitarbeit (other work), which accounts for the other half and is usually equated with the 'oral' aspect of classroom work although its definition includes participation and effort, and practical work as far as applicable, as well.

It is interesting to note that later (§ 22) a careful distinction is drawn between the written classroom tests as referred to above and limited written exercises as a further basis for assessment. These are not prescribed, may be marked, count for the 'other work' half with the value of a weightier oral contribution in class, but are not supposed to replace the assessment of oral skills. More details of this important distinction will be presented below (see section 3.2).

These arrangements imply (i) that assessment is conceived of as a continuous process rather than a once-off event, (ii) that in spite of explicit attempts to safeguard the assessment of oral work there is a noticeable bias towards writing as the basis for assessment, since writing-based procedures account for well over half the weight of the two assessment domains taken together.

While it is appreciated that procedures eliciting and evaluating work samples at several points over the educational year are more likely to yield reliable and valid results than those making the assessment dependant on a single, however massive, examination at the end of a course of instruction, language teachers find that the built-in bias towards writing-based assessment procedures makes it difficult to change the present practice of assessing the oral skills impressionistically, with no prescribed or customary method or format as a guideline, because it is in fact less important for the eventual result at the end of the school year and is considered to be the 'soft' part of the assessment, against the 'hard' writing-based parts. In other words, there is a strong tendency for the results in the written classroom tests - and possibly the marked written exercises - to dominate the evaluation of the achievement of a pupil at the end of the school year, with an independant assessment of oral skills either non-existent or only marginally modifying the assessment predetermined in this way.

3 Assessment of English (Forms 5 - 10)

3.1 Klassenarbeiten (Written Classroom Tests)

3.1.1 Organisatory Conditions

The importance of the written classroom tests for the assessment in general is suitably underlined by the range of rules and regulations that surround them.

As TABLE 1 shows, secondary education starts in the fifth form, which is the first year of English, and can be followed either in one of the school types of the traditional tripartite system, i.e. Hauptschule (Secondary Modern School), Realschule (Secondary Technical School) or Gymnasium (Grammar School), or in a Gesamtschule (Comprehensive School). Compulsory schooling ends after the 10th form. In all schools, English is compulsory. The table sets out how many lessons per week are held in English as a school subject, in each of the school types, how many of the prescribed written classroom tests are to be taken per educational year, and how long these tests are supposed to be (in periods per test).

Table 1

Written Classroom Tests in English (Forms 5 - 10)

year of English	form	"Hauptschule" (Sec.Mod.School) lessons per week	tests per ed. year	length of test in per.	"Realschule" (Sec.Techn. School) lessons per week	tests per ed. year	length (periods)	"Gymnasium" (Grammar School) lessons per week	tests per ed. year	length (periods)	"Gesamtschule" (Compreh. School)
1	5	5	3	1	5	8	1	5	10	1	5
2	6	5	6	1	5	8	1	5	10	1	5
3	7	4	6	1	4	8	1	4	8	1	4
4	8	4	6	1	4	8	1	4	8	1(-2)	4
5	9	3	6	1	3	6	1-2	3	6	1-2	3
6	10	3	6	1	3	6	1-2	3	6	(1-)2	3

(no data available)

It is interesting to note that although the number of English lessons per week is the same in all school types, in the more demanding schools more and longer written classroom tests tend to be required, a perhaps not quite convincing connection made by the system between frequency and length of measurement, on the one hand, and difficulty on the other.

All written classroom tests - as all other assessment procedures within the school system, with the possible exception of the Abitur (see section 6 below) - are set and marked by the teacher of the class. There is an explicit rule (Allgemeine Schulordnung § 22) obliging teachers to adjust the difficulty level of the individual test to the avarage standard of the class. This is implemented through the requirement that in the case of more than one third non-passes the test is to be repeated unless the headmaster of the school is prepared to sanction it, as an exception.

These arrangements are obviously designed to avoid any potential gap that might develop between the teaching and the testing of it: there is no problem for teachers to make sure that their pupils are tested on what they have been taught. It is equally obvious that the arrangements are intended to ensure that teachers do not get away with teaching without proper regard of what their pupils can learn: if more than a third of the pupils have not been able to follow to the degree of at least deserving the lowest pass mark for their efforts in the corresponding test, then the teaching was perhaps not as successful as it should have been, and an obligation to attempt to reach more members of the class seems reasonable. In this context, the 'exception' rule as outlined above apparently functions as a safeguard against possible misuse of the general 'one third' regulation by classes or groups wilfully pursuing a strategy of non-cooperation, or similar.

It is not difficult to sympathise with the pedagogical ethos that seems to have found expression in the regulations described so far. Unfortunately, these arrangements tend to have rather problematic side-effects in that they encourage a form of group-oriented norm-referencing - with each class functioning as an independant group - which amounts to a rather extreme version of what is of course in principle inherent in any system asking teachers to both set and mark their own tests, with few outside reference-points except the individual teacher's ability and conscience. It is temptingly easy for a teacher, when the results in a particular classroom test threaten to fall under the 'one third' rule, to simply raise the general level of the marking a little, rather than to go through the time-consuming process of setting and marking another test on the same field at a later stage: as long as the marking is internally consistent within the group, there is no immediate outside reason why this should not be done.

The consequence is that given marks are extremely difficult to compare across the system, as there are no established prodecures of taking the different averages of the individual groups into account. Marks are therefore (more or less) useful indicators of the rank of

a learner in relation to his own group, but are practically uninterpretable as indicators of achievement across classes, schools, or regions: what is average in one group may be well above or below average in some other group of (formally) the same level, a natural consequence of accidental group composition and aggravated by systematic variations in the marking habits of teachers. These, incidentally, have reputations as 'hard' or 'soft' markers within their schools so that people (i.e. colleagues, pupils, parents) tend to appreciate a good mark from teacher X as equivalent to an average mark from teacher Y, etc.

There are two additional sets of organisatory rules connected with the written classroom tests which are of interest here (see Allgemeine Schulordnung § 22). To avoid surprise attacks, the dates of written classroom tests are to be announced in advance, they are to be marked, then returned to, and discussed with, the pupils (see also next section below), then given to the pupils to take home for their parents to see, and finally to be returned to the school. Further, they are to be spread evenly over the educational year, pupils are not to be given more than one on any single day, and there are to be no more than two in any week.

While both sets of rules make obvious pedagogical sense, the second one is rather important in contributing to the pressure to adjust the marking of a test, in the way outlined above, to avoid a repetition. Since pupils have to take written classroom tests not only in one subject, but in many others besides English, test dates are usually co-ordinated by the various teachers of a class and fixed well in advance, in order to meet the above organisatory requirements. Any later change threatens to overthrow the intricate balance of the various dates for the required number of tests in the different subjects, and the obvious result is that there is a strong interest in avoiding the repetition of a test altogether, as far as possible.

3.1.2 Formats

After the foregoing outline of the most important organisatory requirements connected with the prescribed written classroom tests it is now appropriate to look more closely into the nature of the formats used for this purpose.

Again, we can start from what is officially laid down, in this case in the Richtlinien (guidelines) of which there is a set for each of the subjects taught at school and which are issued by the Ministry of Education.

Taking the Richtlinien for the Grammar Schools as our main example (Kultusminister NRW ed. 1978b), the six years from the 5th to the 10th form are grouped into three two-year bands, with the following examples or recommendations for the written classroom tests:

(1) forms 5/6: answer questions, form questions, describe picture, verbalise picture story, give simple explanations, change dialogue into narrative or vice versa, retell a text changing

the point of view (e.g. from 'He ...' to 'I ...'). A few techniques are explicitly not recommended, in particular dictations (neither of known texts, nor of slightly changed ones, nor of entirely unknown ones), the rendering of a text learnt by heart, the summary of a text (or part of it) read in class (cf. Kultusminister NRW ed. 1978b: 19).

(ii) forms 7/8: all of the above at a more demanding level, plus e.g. characterizations of people, objects, events; summaries (including structural analysis, main ideas, comments on sequence of events); analyses (at a fairly elementary level, at this stage) of stylistic/linguistic properties of given text types; changes from one text type (e.g. letter) to another (e.g. interview with letter writer); production tasks such as 'Describe the school day of a British pupil' (after treating the topic of school life in Britain in class over a period of time) (cf. op.cit. 37-38). There is an explicit note saying that integrative formats should definitely predominate, at this stage (cf. op.cit. 32).

(iii) forms 9/10: all of what is characteristic of the above group, only at a more demanding level and with particular emphasis on the analysis of fictional and non-fictional texts, e.g. work out the structure of the text, the pros and cons, plot, characters, historic background, comment on style/register etc. (cf. op.cit. 55). Again, explicit reference is made to the need of using integrative formats which alone can do justice to the complex objectives of the teaching at this level. This is considered to be more important than the loss in objectivity which this approach inevitably entails: "Die mit diesen Formen verbundene verminderte Objektivität der Leistungsbewertung muß mit in Kauf genommen werden." (op.cit. 56).

An attempt to see these guidelines in perspective reveals two interesting tendencies. The first is that written classroom work is clearly intended to set the model for the assessment formats employed to monitor the learning progress. In other words, the written classroom tests are to elicit work samples representative of the written activities performed in class, so that there is in fact an identity between exercise and assessment formats. Teachers tend to be rather in favour of such a scheme, to be sure for good pedagogical reasons.

The second tendency is less apparent in the letter than in the spirit of the guidelines, which implicitly favour a criterion-referenced approach - by outlining learning objectives and recommending assessment procedures asking for judgments as to whether and how far these objectives have been reached by the learners - in spite of, and in contrast to, the strong norm-referencing which is suggested by the external framework of the assessment system in general (cf. previous section). Whether these two contradictory approaches can be reconciled in a meaningful way, whether they cancel each other out to lose the advantages of them both, or whether this simply leads to a situation where anything goes and no obligation is felt towards any of

these principles, thus leaving the exact purpose of the assessment conveniently in the dark, is a question on which little information is available to date and which can only be raised rather than gone into here.

Practical realisations of the recommendations of the guidelines can be found in the Richtlinien themselves, which frequently draw attention to assessment formats when discussing exercises considered helpful in achieving the learning objectives. In addition, there are a number of useful publications available as help for the teacher (e.g. Schulz et al. 1984) which provide detailed discussions of concrete examples and often include marking and grading suggestions as well as handy bibliographies (2).

3.2 Tests (Marked Written Exercises)

The emphasis on integrative formats as described above does not mean that discrete item formats such as series of multiple choice, true-false, matching, fill-in or similar items, which are usually associated with objective testing, are not acceptable within the system of assessment at all. But they are considered mainly suitable for checking off some of the more isolated, mechanical aspects of language and are thus given a more limited rôle: tests consisting of tasks with these formats fall outside the regulations governing the prescribed written classroom tests, they are non-compulsory, and their results count towards the 'other work' area (cf. section 1 above; cf. also Kultusminister NRW ed. 1978b: 11).

Within the context of the whole system, this distinction, which is no doubt quite rational as such, unfortunately encourages two types of misunderstandings which contribute to making the 'other work' half of the assessment a rather problematic field.

The first is that in spite of explicit exhortations to the contrary (cf. Allgemeine Schulordnung § 22) discrete item tests, with their status of (non-compulsory) marked written exercises, are easily confused with a 'hard' way of measuring oral competence. In fact, even the Richtlinien themselves, in this case the ones for English in Comprehensive Schools, explicitly make this very assumption in a somewhat surprising statement suggesting that observing and assessing the oral command of the language can very well be carried out by informal written tests:

> Es wird ... häufig notwendig sein, mündliche Fertigkeiten wie Sprechen aus Gründen der gesicherten und schnelleren Bewertbarkeit durch schriftliche Kontrollverfahren zu überprüfen. (Kultusminister NRW ed. 1980: 142).

The second misunderstanding connected with discrete item tests within the present system is that they are in fact not quite so 'hard', i.e. objective, measures as one might think, because they are commonly implemented as quickly assembled lists of individual items, with little attention to the technique employed and the validity and reliability of the items included. In keeping with the tendencies of

the system in general, they are teacher-made and used individually for a particular group at a particular time. No pre-testing is done, nor are there any other steps taken to ensure that they can at least develop into anything more than isolated idiosyncratic snapshots of which language items the learners happen to 'know'.

Although these marked written exercises are commonly called Tests in the German school system, in contrast to the prescribed Klassenarbeiten (written classroom tests) described in the preceding sections, they are in fact rather remote from tests in the real sense of the word.

While teachers, understandably enough, often do not have the expertise, nor in particular usually the time, to improve the test qualities of these home-made 'tests' themselves, there is a somewhat surprising reluctance to use e.g. existing standardised tests, where after all this work has already been done for them. One may speculate that these are perhaps not sufficiently well known; it could also be that differences between various groups are so considerable that these tests are vaguely felt to be too remote from what is actually being taught in this particular classroom.

The alternative could be to develop the 'tests' into a network of criterion-referenced assessments, as another way of gaining a reference point outside the class for at least part of the total of the assessment procedures used. Again, it would be possible to draw on published tests of this nature, but again teachers appear reluctant to use them, possible for similar reasons as above.

The impression one gathers from an analysis of the Tests (in the curious understanding of the term given to it in the German school system) is that they are not an area of great interest or concern for teachers who are under enough pressure in other fields not to worry too much about this one. Considering the rather limited weight 'tests' have, as no more than non-compulsory parts of one half of the total assessment, this neglect is perhaps not totally surprising.

4 Correcting and Marking of Klassenarbeiten

What is accepted as more important, on the other hand, is the need for some kind of uniformity in the correcting and marking of the - more important - written classroom tests. In order to gain some insight into the problems connected with this field, it might be interesting to see which suggestions were made in a recent in-service training meeting (Landesinstitut für Schule und Weiterbildung November 1984), and to add some comments on them.

The problem of awarding an appropriate mark within the range of the marking scale prescribed by the system is only vaguely touched on by a reference to the learning and teaching objectives as laid down in the official guidelines (Richtlinien) already referred to previously: inhowfar these have been reached in a given written classroom test (Klassenarbeit) remains entirely a matter for the individual teacher to decide on and to convert into a mark. No guidelines are offered for

this most problematic aspect of any assessment. Guidelines are, however, considered more appropriate in connection with the correction and marking of mistakes.

All mistakes are to be marked and classified as to kind and gravity. The following categories are to be used and indicated by the appropriate abbreviation: Gr - Grammatik (grammar), T - Tempus (tense), Bz - Beziehung (concord), Sb - Satzbau (syntax), St - Stellung (word order), Pr - Präposition (preposition), W - Wort (lexical item), A - Ausdruck (collocation, expressions of more than one word), R - Rechtschreibung (orthography), Z - Zeichensetzung (punctuation). All of these count as one mistake, except for the last two which count only as 0.5 error unit. However, in dictations - which are considered acceptable only (if at all) at a rate of no more than one per half year in forms 5 and 6 - orthographic mistakes also count as one whole error unit. There is an explicit reminder that only those imperfections in the pupils' papers are to be counted as mistakes which can be shown (e.g. by dictionary, grammar book, reference book) to be definitely wrong: it is not enough for them to just 'sound wrong' to the teacher.

A number of additional specifications provide rulings for individual problems. There is a maximum of one mistake per word. A - mistakes, which by definition extend over more than one word, count as one single error unit, but other mistakes (e.g. orthography, grammar) count in addition, e.g. 'He locked after the key' would have one A and 0.5 R mistakes. Orthography counts as one mistake if the result is a different word, e.g. 'there holidays'. Similarly, if orthography is connected with grammar it counts as a (full) G - mistake, e.g. 'in it's nest'. Recurring mistakes are marked but counted only once; however, tense errors count each time unless there is an internally consistent tense shift; recurring elementary errors (e.g. in connection with the 3rd person singular -s) are treated as other recurring mistakes but if frequent are to be taken into consideration when determining the final mark.

After the correcting and marking, the written classroom tests are to be returned to the pupils in class. Teachers are expected to discuss the tests so that the pupils (i) understand why they have been given the mark they have received and (ii) can correct their own tests. The explicit idea is to help pupils to avoid making the same mistakes in future.

It can be assumed that similar suggestions are in principle in operation in many areas and countries. What is perhaps not quite satisfactory is that many of the error categories overlap considerably, so that the clarity suggested by the abbreviations is more apparent than real. Considering the orthography problems even of native speakers of English, the recategorisation of some orthographic mistakes as lexical (there - their) or grammatical mistakes (its - it's) is not entirely convincing. And inhowfar discussing the tests in class when returning them to the pupils contributes to helping them to avoid making the same mistakes in future is in fact rather a thorny question: there is little doubt that more systematic follow-up work than is pos-

sible within the one period usually devoted to it would be needed for this aspect of classroom work if the procedure is to be more than a ritual.

5 Assessment in English (Forms 11 - 13)

In the above outline of the ways in which achievement in English is measurement within the German school system an attempt has been made to show in particular the general pattern of assessment during the six years of compulsory secondary schooling, with its emphasis on the written classroom tests, the problematic evaluation of the 'other work' area, and the curious rôle of the marked written exercise. When turning now to the final three years of instruction in English offered by the system to pupils attending Grammar Schools (and some Comprehensives) it will be sufficient to highlight the differences to this pattern, before finally looking in some detail at what is demanded in English in the Abitur, i.e. the final pre-university level school-leaving examination.

According to the Richtlinien for forms 11 - 13 (see Kultusminister NRW ed. 1981), the assessment of English at this level is intended to probe into three areas, (i) the command of the language as such, (ii) knowledge about e.g. language (→ linguistics), texts (→ literature), socio-cultural backgrounds (→ history, 'civilisation'), and (iii) mastery of methods (e.g. how to deal with an 'unseen' text, how to use a monolingual dictionary, a reference book etc.). These domains are not to be elicited and assessed independently of each other, but learners are to be given the opportunity of demonstrating their abilities in all three of them.

While the way in which this is envisaged will be sketched out below it is important to note at this stage that these demands constitute a definite shift and widening of the scope of assessment to include, above and beyond the foreign language as such, a strong content component (if we leave the 'method' aspect out of consideration for a moment), as a legitimate part of the total evaluation: in order to be successful in English it is now no longer enough to have an acceptable command of the language, but it is also important to be able to cope with the information retrieval and processing as it is required by the contents to be carried by the foreign language.

Since there is a considerable degree of overlap between what is demanded here and in other arts subjects (such a German as mother tongue, other foreign languages), in particular as far as methods, but also as far as some content areas are concerned (knowledge about language, text types, literary analysis etc.), high correlations between an individual learner's sucess in English and in these subjects can be expected. On the other hand, at a stage where a foreign language has been taught for over six years, language can only be expected to develop further in connection with contents that can be taken seriously and is a real challenge, so that there are many good reasons for broadening also the basis of assessment in this way.

5.1 Sonstige Mitarbeit ('Other Work') Domain

The basic pattern of assessment is still essentially the same as the one described above for the younger year groups, i.e. the assessment of achievement is based on (i) a prescribed number of teacher-made written classroom tests (now called Klausur), and (ii) the Sonstige Mitarbeit ('other work').

Generally, criterion-referencing is emphasized more strongly at this level, with the result that the 'one third' rule (see 3.1.1 above) is in fact somewhat relaxed (cf. Kultusminister NRW ed. 1981: 121-122; cf., however, 169).

There is a strong emphasis on the equal weight of these two assessment domains, underlined by the requirement to explicitly summarize, in a separate mark each, the results of the written classroom tests, on the one hand, and of the 'other work' domain, on the other, before drawing the two together into a single mark for the subject at the end of each half-year (cf. op.cit.: 122). The importance of the second domain of assessment is further emphasized by the fact that for some pupils in the final year, for whom the written classroom tests are under certain circumstances no longer compulsory, it may even constitute the only basis of assessment in the subject (cf.ibd.).

This makes it necessary for the Richtlinien to spell out in some detail how the domain of Sonstige Mitarbeit is to be interpreted (cf. op.cit.: 122-129). A basic distinction is drawn, for this purpose, between continuous assessment procedures on the basis of observation, and more formal assessments of individual pieces of work at individual points.

The first approach is considered to be particularly appropriate for the evaluation of the oral work done in class, where an assessment at one or two individual points only is felt to be too limited to capture a fair sample of a learner's contributions to oral classroom work. There is no doubt that the assessment of this field is intended to be the weightiest part of the 'other work' domain: there is an explicit note saying that - in conjunction with an assessment of the homework (see below) - it may well suffice as the basis for an assessment of the domain as a whole (cf. op.cit.: 124).

However, in spite of a detailed list of criteria that can be taken into consideration when evaluating oral performance at this level (cf. op.cit.: 149-150), teachers and pupils feel rather uneasy about the subjectivity of the impression marking that is practically unavoidable in this field. In a situation where teachers for good pedagogical reasons refuse to rush to their note-books to record a mark every time anyone says anything in class, and where they teach up to six different classes every morning, the only way for the conscientious teacher to arrive at a broad basis for the eventual assessment of the oral performance of his pupils is to systematically record marks from memory after school on the pupils taught that day. Under these circumstances, however, it is already sufficiently problematic to accurately recall who said what, to make a systematic application

of the criteria suggested practically illusory.

The second approach to the evaluation of the Sonstige Mitarbeit domain consists in the assessment of pieces of work elicited at individual points over the school year. Four types of work are recommended as particularly suitable, in a subject such as English, for this purpose (op.cit.: 124-129), viz. (i) homework, which at this level is considered to be complementary to classroom work and can e.g. also be of a preparatory nature, (ii) the minutes of a lesson, i.e. a somewhat detailed summary of the main points covered, (iii) marked written exercises (on which more below), and (iv) prepared talks on special, individually selected topics. Of these, only the marked written exercises lead to a direct comparison of all members of the same group; all the other types are usually interpreted as pieces of work which individual learners may be asked to do and hand in and/or read out in class, to be given a mark for individually.

This constellation tends to put the marked written exercises into a more prominent position. Consequently, a number of rules are in force to define their rôle more precisely and determine the organisatory framework in which they are to be used. As far as the latter is concerned, there are to be no more than two of them per half year. No more than two marked written exercises in different subjects may be taken by a pupil on any one day; when a pupil is due to take a written classroom test (in any subject) no marked written exercise is allowed for this pupil at all on the same day. Advance notice must be given, and the tasks set must allow completion within about 30 (and an absolute maximum of 45) minutes.

The marked written exercises are intended to provide insights into the ways in which pupils handle a particular limited though comprehensive question or problem connected with what is being taught at the moment. Series of short independent questions are explicitly ruled out, as are e.g. vocabulary tests (op.cit.: 127). Although in practice these latter stipulations are perhaps the regulations most frequently ignored in the whole context of assessment it is quite clear that the marked written exercises are meant to be short versions of the written classroom tests. Interpretations e.g. in terms of discrete item formats, which are still conceivable for the corresponding purpose in the lower forms (cf. 3.2 above), are not in accordance with what is intended, at this level.

5.2 Klausuren (Written Classroom Tests)

As pointed out above, the assessment of achievement is based, in equal parts, on the 'other work' domain, some details of which have been highlighted in the preceding paragraphs, and the prescribed written classroom tests, to which we shall turn our attention now.

The organisatory arrangements for this domain are in principle comparable to the ones for the lower forms (cf. 3.1 above). TABLE 2 shows that at this level English can be taken as an ordinary subject (Grundkurs) at three lessons per week, or as one of the two major

TABLE 2

Written Classroom Tests in English (Forms 11 - 13)

year in English	form	taken as ordinary subject (Grundkurs): lessons per week	tests per half year	length of test (periods)	taken as major subject (Leistungskurs): lessons per week	tests per half year	length of test (periods)
7	11.I	3	2	2-3	6	2	2-3
	11.II	3	2	2-3	6	2	2-4
8	12.I	3	2	2-3	6	2	3-5
	12.II	3	2	2-3	6	2	3-5
9	13.I	3	2	2-4	6	2	4-6
	13.II	3	1	3 hrs.	6	1	5 hrs.

subjects (Leistungskurs) required of every pupil in the final years, at 6 periods per week. The school year is divided into two halves (marked as I and II in the table), with in some cases slightly different regulations each. The final written classroom test in 13.II is given in hours instead of periods, as an approximation to the Abitur regulations (see section 6 below).

Learners are not to take more than one of these in any subject on any one day, and no more than a maximum of three per week. We have mentioned before that under certain circumstances some pupils are not required to take the written classroom tests in the final year at all: these pupils are not allowed to take them even if they wish - which they might want to do feeling happier with this kind of written work than having their assessment based entirely on the 'other work' domain - nor is it possible for them to be given a marked written exercise while the others do the Klausur. An interesting note stipulates that a written classroom test must have been handed back to, and discussed with, the pupils at least a day before the next one can be given (cf.

op.cit.: 169), as a - perhaps not entirely unnecessary - reminder to teachers of the pedagogical, in addition to the measuring, purpose of them.

As learners at this stage in their school careers are no longer taught in stable groups identical for most subjects but have in fact individual timetables regrouping them for practically all the subjects they take, a lot of organisatory dove-tailing - usually computerised these days - is required to meet the above conditions. Dates for the written classroom tests are worked out by the school administration, instead of the teachers, and are consequently felt to be prescribed from outside in a much stronger sense than in the lower forms. There is no doubt that pedagogical considerations, such as giving a test when the group is ready for it rather than two weeks before that stage or three weeks after starting on a new area already, plays no part in fixing the dates, on which teachers have practically no influence at all any more.

In keeping with the broadening of the assessment to include a strong content component as well, the evaluation of a written classroom test is based on an assessment (i) of the content, where errors would result from factual inaccuracies or logical inconsistencies etc., and (ii) of the language, where errors would result either from problems with formal aspects of the language (with correction rules similar to those outlined in section 4 above for the lower forms; cf. op.cit.: 168-169) or from problems with what is called 'text production' which is interpreted as the overall composition of the text produced by the pupil (cf. op.cit.: 167). It is recognised that language problems cannot always be attributed clearly to one or the other of the two language fields, but teachers are advised to take care that individual errors are counted only in one of them (op.cit.: 169). The overall mark awarded for a written classroom test is the result of a separate assessment of the content and the language, with the evaluation of the language as a rule the weightier factor (cf. op.cit.: 169).

The most problematic part of these regulations is perhaps the separation of the language part of the assessment into a formal aspect on the one hand and a style/composition aspect on the other. This is obviously intended to make it possible to strike a balance between these two - often non-correlating - aspects of a learner's written effort. However, the problems of operationalising the style/composition aspect into anything more than impression marking, or inspired guesswork, are considerable. Thus, it is perhaps not uncharacteristic of the situation that the overlap with the content area (when it comes to an assessment of the 'composition', i.e. the effectiveness of the presentation of the points made) which is after all parallel to the recognised overlap at the other end with aspects of the language form (in particular concerning lexical items), seems to have escaped the Richtlinien's notice. In practice, therefore, the evaluation of the style/composition aspect is usually considered to be the 'softest' part of the overall assessment which can be used to justify the mark which a test appears to deserve in the first place (3).

BEST COPY AVAILABL

6 Abitur - Assessment in English

To round off our survey of the ways in which achievement in English is measured within the German school system, we can finally focus briefly on the arrangements made in connection with the pre-university level school-leaving examination (Abitur), for this purpose.

In accordance with the general Abitur regulations, the assessment in English, as it eventually appears on the certificate of the successful candidate in form of a mark, is in fact partly based on the results achieved within the final two years in the subject, and only in part on those obtained in the final examination itself, at a proportion of 2 : 1.

The final examination, on its part, can either take the form of a written test similar in type to the Klausuren described in the previous section, with an additional oral examination only if the results in the written part are markedly different from the average achieved throughout the final two years (or if the candidate hopes to improve his/her mark by taking one), or it can take the form of an oral examination alone in cases where candidates meet the required number of written tests in other subjects. While the complicated details of which candidates take which form of the examination are perhaps irrelevant here (for details see e.g. Kultusminister NRW ed. 1982), a short survey of aims, formats and evaluation procedures for both the written and the oral parts may be instructive (cf. Kultusminister ed. 1981: 169-235).

Similar to the Klausur regulations, the written examination is to give candidates the opportunity of demonstrating their command of the language, of particular fields of knowledge, and of adequate working methods.

To ensure some kind of conformity in the kinds of tasks set at this level, an agreement was reached in the Federal Republic in 1980 (cf. Sekretariat der Ständigen Konferenz der Kultusminister ed. 1981) to restrict the possible formats to the following four types:

(i) The candidate is given a text (fiction) of about 700 words (major subject; time allowed: 5 hours) or 500 words (ordinary subject; time allowed: 3 hours) (± 100 words) plus 4 - 6 questions specifying what is to be done with it, touching on content, form and an evaluation of the text and leading from comprehension to production.

(ii) The candidate is given a non-fictional text, with tasks specified as under (i).

(iii) The candidate is given a combined task 'Type A', which consists of

(a) a listening comprehension passage of 300 - 600 words (maximum of 4 minutes on tape) with 6 - 8 narrow questions eliciting

major content points of, and comments on, the text. The procedure is narrowly prescribed: (1) listen, (2) look at the questions, (3) listen again, (4) answer questions (time allowed: 30 minutes), (5) hand in. Only then is it possible to go on to the next part,

(b) a text-based task (fiction or non-fiction) similar in type to what is demanded under (i) or (ii) but 'shortened accordingly' (cf. Kultusminister ed. 1981: 177).

(iv) The candidate is given a combined task 'Type B', which consists of

(a) a picture, photo, cartoon or similar, accompanied by 3 - 4 questions eliciting a description and interpretation (maximum time allowed no more than one third of the total), and

(b) a text-based task similar to (iii .b).

In keeping with the tendencies of the whole system, it is the teacher of the candidates concerned who sets the tasks, which, however, in order to conform to some kind of common standard, need to be approved of by the regional Board of Education. The procedure is for teachers to prepare two packages of two different formats each and to hand these in to the appropriate Board of Education who select one of these packages (or reject asking for new suggestions) for presentation to the candidate who eventually chooses one of the two tasks included in it (4).

The criteria for the assessment of a candidate's work are in principle comparable to those for the Klausuren (cf. op.cit.: 184-187). Marking is done by the teachers who set the examination, with another teacher of the same school (usually appointed by the headmaster) acting as second examiner and the Board of Education exercising its power of inspection at its own discretion.

Of the four formats theoretically allowed, only the first two are in regular use, with the other two in practice occurring only as rare exceptions. This is of course not really surprising: the listening comprehension and 'viewing comprehension' sections of the combined tasks cannot build on any systematic preparation in previous assessment contexts and are therefore avoided as uncalculable risks; they are in addition surrounded by so many restrictive regulations (possibly originally designed to keep these two formats from developing into 'soft options') that they are technically unnecessarily difficult to handle; and the text-based formats are the most important parts in any of the four versions anyway, in the first two accounting for the whole, but in the last two still dominating the assessment as the weightiest factor by far, so that concentration on text-based formats appears rather more relevant - and sufficient - as preparation for the final examination (5).

The oral examination in English, as far as it is carried out (see above), also has a narrowly prescribed format. The aims are

similar to those of the written test, and the task is also text-based in that the candidate receives a text of about 300 words with 2 - 3 specifications eliciting aspects of content, form and comment from the candidate, who is allowed 30 minutes' preparation. The examination itself takes 20 - 30 minutes, the candidate is supposed to produce answers to the first two specifications as a continuous presentation with the third specification treated in form of an interview, i.e. by question and answer.

The oral examination is conducted by a small examination committee consisting of a chairman and three other members, one being the examining teacher, one acting as assessor, and one keeping the minutes. The examiner is required to prepare the tasks and pass them on to the examination committee at least two days in advance of the examination, for approval. It is the chairman's responsibility to ensure that all relevant regulations have been adhered to. If an inspector from the Board of Education is present, he would usually chair (some of) the orals.

It is interesting to note that in the oral examination listening comprehension tasks are explicitly excluded, if not entirely for the second part, then at least for the first (cf. op.cit.: 217). Another curious feature is the fact that minutes are to be kept which are to mirror rather precisely the most important aspects of the examination, including characteristic examples of the strengths and weaknesses of a candidate's language performance, but recordings are not permitted.

7 Conclusions

In the present paper an attempt has been made to provide an outline of the ways in which achievement in a foreign language such as English is assessed within the context of the German school system. After a brief exposition of the general organisatory framework of all assessment procedures at school, we have tried to describe the situation for the forms 5 - 10 and 11 - 13 of secondary education, as two different groups, before turning to the regulations that govern the conditions for the Abitur examination. It is hoped that the outline as presented here has provided a fair picture of the situation. Perhaps unavoidably, problematic aspects have been discussed in more detail than others. To redress the balance, attention should perhaps be drawn again to the great flexibility and freedom of a system which allows its teachers to tailor their assessment procedures to fit their teaching as closely as is conceivably possible, for most assessment purposes within the nine years in which English is taught at school. In addition, there seems to be an untapped potential for criterion-referenced assessment not sufficiently exploited so far.

On the other hand, there is little doubt that the extreme classroom-oriented norm-referencing demanded by the 'one third' rule encourages a situation in which outside criteria - i.e. questions such as 'How does a learner stand, and what can he do with his English, outside this particular classroom?' - unfortunately appear to be to-

tally irrelevant. This does, indeed, raise the questions of accountability in a rather serious sense.

As any other system that has grown and been modified in individual points over a longer period of time, the system described has its fair share of inconsistencies and problem areas, e.g. the assessment of the Sonstige Mitarbeit domain, the marked written exercises, the dominance of the text-based procedures with their backwash effect on the teaching, and many others. However, not all formats and procedures are problematic in themselves: if one were to pinpoint the single most effective measure to make the system more compatible with what professional language testers would have to demand, it would be a move to re-organise the marking. Rater training, multiple marking and independent marking would go a long way towards remedying a situation which in many ways in its present state of unsophistication would have to be called 'pre-scientific' in Spolsky's sense of the word (Spolsky 1975).

Notes

1 For a possible distinction between e.g. 'tests' and 'examinations' cf. Britsh Council ed. 1976.

2 Examples of tests actually used for these year-groups were made available to the participants of the conference through the courtesy of Mrs Irmgard Voss.

3 A range of publications is available to teachers seeking suggestions, examples and guidelines on the problems of setting and marking a Klausur, e.g. Bliesener 1981a, 1982, Hurst et al.1980, Multhaupt 1983 (for comprehensive bibliography). Examples of tests actually used recently were again made available to participants of the conference through the courtesy of Mrs Voss.

4 Again, examples were made available through courtesy of Mrs Voss.

5 For publications on the Abitur regulations (examples, discussions, criticism etc.) see relevant titles in the bibliography.

Bibliography

Ahrens, R., and W. Hüllen. 1977. "Bemerkungen über die 'Einheitlichen Prüfungsanforderungen in der Abiturprüfung: Englisch'". Neusprachliche Mitteilungen: 25 - 29.

Bliesener, U. 1981a. Klausur- und Abituraufgaben Englisch. München: Langenscheidt-Longman.

- 1981b. "Die neue Abituraufgabe Englisch gemäß den Einheitlichen Prüfungsanforderungen. Überlegungen zum sinnvollen Umgang mit

der Textaufgabe". Praxis 28: 227 - 238.

- 1981c. "Arbeitsanweisungen im schriftlichen Abitur - Englisch". Praxis 28: 357 - 367.

- 1982. Klausuren und Abiturarbeiten in Englisch. München: Langenscheidt-Longman.

Bludau, H. 1976. "Normenbuch Englisch". Neusprachliche Mitteilungen: 68 - 70.

The British Council (ed.). 1976. Examinations and Tests in English for Speakers of Other Languages. Information Guide No. 6.

Hurst, N., Nissen, R., and U. Schulz. ^{2}1980. Oberstufenarbeiten Englisch. Von der Textauswahl bis zur Bewertung. Dortmund: Lensing.

Kornelius, J., and E. Otto (eds.). 1983. Anglistik. Richtlinien und Englischunterricht. Trier: Wissenschaftlicher Verlag Trier.

Der Kultusminister des Landes Nordrhein-Westfalen (MSW) (ed.). 1978a. Allgemeine Schulordnung (ASchO) vom 8.11.1978. In Der Kultusminister (ed.) ^{2}1979: 107 - 135.

- 1978b. Gymnasium Sekundarstufe I: Vorläufige Richtlinien Englisch. Köln: Greven.

- ^{2}1979. Handbuch Schulmitwirkung. Köln: Greven.

- 1980. Gesamtschule: Richtlinien Englisch. Köln: Greven.

- 1981. Gymnasiale Oberstufe: Richtlinien Englisch. Köln: Greven.

- 1982. Die gymnasiale Oberstufe. Informationsschrift für Schüler. Bochum: Kamp.

Multhaupt, U. 1983. "Nachgetragene und kommentierte Bibliographie zu den Richtlinien für die gymnasiale Oberstufe in Nordrhein-Westfalen, Englisch". In Kornelius and Otto (eds.). 1983: 200 - 225.

Landesinstitut für Schule und Weiterbildung Nordrhein-Westfalen. 1984. "Fachübliche Grundsätze für die Korrektur und Bewertung von englischen Klassenarbeiten" mimeo.

Pürschel, H. 1981. "Die Einschätzung und Bewertung mündlicher Leistungen durch deutsche Englischlehrer". Praxis 28: 368 - 375.

Schulz, R., Eckardt, I., Sammler, M., and D. Williams. 1984. Klassenarbeiten im Englischunterricht der Klassen 7 - 10. Beispiel Realschule. Dortmund: Lensing.

Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (ed.). 1981. Einheitliche Prüfungsanforderungen in der Abiturprüfung - Englisch. Beschluß